

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет електроніки
Кафедра звукотехніки та реєстрації інформації

НАУКОВО-ТЕХНИЧНА КОНФЕРЕНЦІЯ СТУДЕНТІВ

***Сучасні проблеми застосування електронних
та інформаційних технологій в телекомунікаціях,
телебаченні та цифровому кінематографі***

25 травня 2017 р.

КИЇВ

Секція В МЕТОДИ ТА ЗАСОБИ СТВОРЕННЯ ТА ПЕРЕТВОРЕННЯ АУДІО ТА ВІДЕО КОНТЕНТУ

**Керівник к.т.н., доцент Трапезон К.О.
Секретар старший викладач Гумен Т.Ф.**

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ПО ЗВУКУ

Баран В.С.

КПІ ім. Ігоря Сікорського, кафедра звукотехніки та реєстрації інформації

Останнім часом нейронні мережі досягли чималого прогресу в області розпізнавання об'єктів і сцен на відео. Дивлячись фотографії чи відеоролики, комп'ютер може практично безпомилково визначити місце дії. Проте, в області розпізнавання звуків нейронні мережі поки що не продемонстрували такого ж прогресу. Спеціалісти із Лабораторії інформатики і штучного інтелекту (CSAIL) Массачусетського технологічного інституту розробили систему машинного навчання SoundNet, яка вирішує цю проблему.

Можливість визначення місця дії по звуку – не менш важлива задача, ніж визначення місця по відео. Оскільки картинка з камери може бути розмитою, вона не буде давати достатньої інформації. Але якщо працює мікрофон – робот зможе зорієнтуватися, де він знаходиться.

Щоб навчити нейронну мережу SoundNet, працівники CSAIL використали метод природної синхронізації між машинним зором і машинним слухом, навчивши нейромережу автоматично добувати звукову репрезентацію об'єкта з нерозміченого відеоматеріалу. Для навчання використали близько 2 млн. відеороликів сервісу Flickr (26 ТБ даних), а також базу анотованих звуків – 50 категорій і приблизно 2000 зразків.

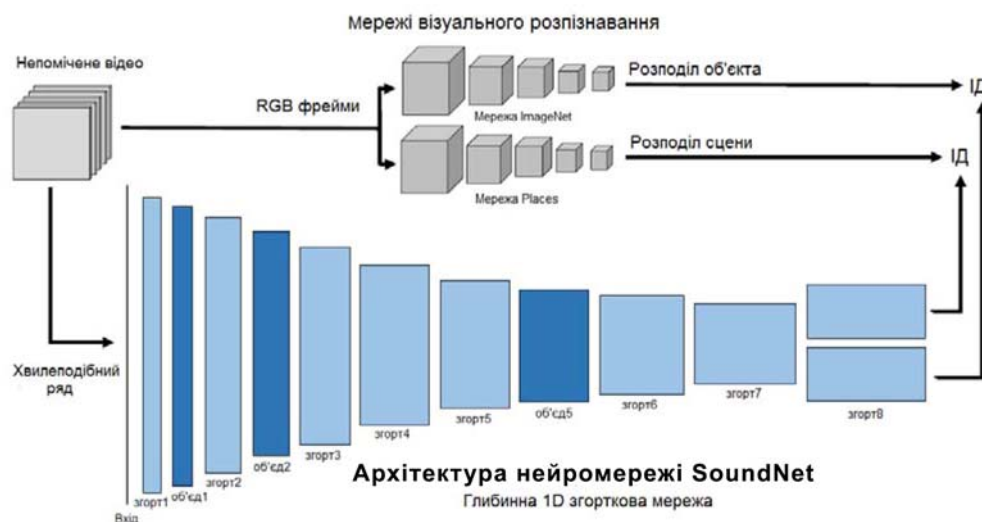


Рисунок 1 – Архітектура нейромережі SoundNet

Хоча навчання нейромережі відбувалося під візуальним наглядом, система видає чудовий результат і в автономному режимі по класифікації мінімум трьох стандартних акустичних сцен, по яким її перевіряли розробники. Більше того, перевірка нейронної мережі показала, що вона самостійно навчилася розпізнавати характерні для деяких сцен звуки, хоча розробники не надавали їй зразків для розпізнавання саме цих об'єктів. Одночасно зі сценою вона розрізняє і конкретний об'єкт, який є джерелом звуку.

Перевірка SoundNet здійснювалась на двох стандартних базах звукових записів, і показала на 13-15% більш високу точність розпізнавання об'єктів, ніж краща із подібних програм. На наборі даних із 10 різними категоріями звуків SoundNet класифікує звуки з точністю 92%, а на наборі даних з 50 категоріями – 74%. На цих самих наборах даних, люди показують точність розпізнавання, в середньому, 96% і 81% відповідно.

У майбутньому такі комп'ютерні програми знайдуть прикладне практичне застосування. Орієнтування по місцевості за допомогою звуку допоможе в програмах управління для автономних роботів та інших машин. У системах безпеки і розумних домах система може автоматично реагувати на конкретні звуки. Наприклад, на звук розбитого вікна. Також технологія допоможе боротися зі звуковим забрудненням вулиць міста.

Перелік посилань:

1. <https://geektimes.ru/post/283332/>.
2. <https://arxiv.org/abs/1610.09001>.

Науковий керівник старший викладач Гумен Т.Ф.